A complete algorithm to resolve ambiguity for intersubunit NOE assignment in structure determination of symmetric homo-oligomers

Shobha Potluri, Anthony K. Yan, Bruce R. Donald and Chris Bailey-Kellogg

Protein Sci. 2007 16: 69-81
Access the most recent version at doi:10.1110/ps.062427307

A complete algorithm to resolve ambiguity for intersubunit NOE assignment in structure determination of symmetric homo-oligomers

SHOBHA POTLURI,[1] ANTHONY K. YAN,[2] BRUCE R. DONALD,[2,3] AND CHRIS BAILEY-KELLOGG[1]

[1]Department of Computer Science, Dartmouth College, Hanover, New Hampshire 03755, USA
[2]Department of Computer Science, Duke University, Durham, North Carolina 27708, USA
[3]Department of Biochemistry, School of Medicine, Duke University Medical Center, Durham, North Carolina 27708, USA

(RECEIVED July 5, 2006; FINAL REVISION September 25, 2006; ACCEPTED October 5, 2006)

Abstract

Assignment of nuclear Overhauser effect (NOE) data is a key bottleneck in structure determination by NMR. NOE assignment resolves the ambiguity as to which pair of protons generated the observed NOE peaks, and thus should be restrained in structure determination. In the case of intersubunit NOEs in symmetric homo-oligomers, the ambiguity includes both the identities of the protons within a subunit, and the identities of the subunits to which they belong. This paper develops an algorithm for simultanous intersubunit NOE assignment and C_n symmetric homo-oligomeric structure determinations, given the subunit structure. By using a configuration space framework, our algorithm guarantees completeness, in that it identifies structures representing, to within a user-defined similarity level, every structure consistent with the available data (ambiguous or not). However, while our approach is complete in considering all conformations and assignments, it avoids explicit enumeration of the exponential number of combinations of possible assignments. Our algorithm can draw two types of conclusions not possible under previous methods: (1) that different assignments for an NOE would lead to different structural classes, or (2) that it is not necessary to uniquely assign an NOE, since it would have little impact on structural precision. We demonstrate on two test proteins that our method reduces the average number of possible assignments per NOE by a factor of 2.6 for MinE and 4.2 for CCMP. It results in high structural precision, reducing the average variance in atomic positions by factors of 1.5 and 3.6, respectively.

Keywords: nuclear Overhauser effect (NOE) assignment; nuclear magnetic resonance (NMR) spectroscopy; protein complex structure determination; homo-oligomer; symmetry; complete search; configuration space; ambiguity; hierarchical subdivision

Symmetric homo-oligomers are an important class of proteins, responsible for functions such as ion transport and cellular regulation. While their structures provide valuable insights into their mechanisms, homo-oligomers are challenging targets for structure determination. In structure determination by nuclear magnetic resonance (NMR) spectroscopy, experimental nuclear Overhauser effect (NOE) data are interpreted as distance restraints between pairs of atoms, and algorithms then compute conformations that satisfy the restraints and are also

Reprint requests to: Chris Bailey-Kellogg, Department of Computer Science, Dartmouth College, Hanover, NH 03755, USA; e-mail: cbk@cs.dartmouth.edu; fax: (603) 646-1672; or Bruce R. Donald, Department of Computer Science, Duke University, Durham, NC 27708, USA; e-mail: brd@cs.duke.edu; fax: (919) 660-6519.

Abbreviations: NOE, nuclear Overhauser effect; NMR, nuclear magnetic resonance; RMSD, root-mean-square deviation; SCS, symmetry configuration space; ACR, ambiguity-consistent regions; WPS, well-packed satisfying; S^2, space of symmetry axis orientations represented on a 2-sphere.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.062427307.

consistent with physical modeling terms. Intra-subunit NOEs are used to determine the structure of a subunit in the symmetric homo-oligomer, whereas intersubunit NOEs are used to determine the complex structure. In this paper, we focus on intersubunit NOEs.

There can be several sources of ambiguity in assigning an intersubunit NOE to a pair of atoms: *intra- versus intersubunit ambiguity*, whether the restrained atoms are within the same subunit in the complex or in different ones; *subunit ambiguity*, the subunits to which the restrained atoms of an intersubunit NOE belong; *atom ambiguity*, the identities of the restrained atoms among those with chemical shifts similar to the NOE peak; and finally ambiguity caused by spurious NOEs (NOEs for which all assignments are wrong). Intra- versus intersubunit ambiguity can be resolved experimentally, by a combined analysis of 3D ^{15}N-edited, double ^{13}C-filtered NOESY of a mixed sample and standard 3D ^{13}C-edited NOESY of a homogeneous ^{15}N, ^{13}C-labeled sample (Walters et al. 1997; Zwahlen et al. 1997). Atom ambiguity can (in principle) be resolved experimentally by using 4D or higher dimensional NOESY spectra (Kay et al. 1990; Fiorito et al. 2006). However, in general, subunit ambiguity, and atom ambiguity for 3D NOESY spectra, require computational solutions. This paper presents the first complete algorithm for resolving subunit and atom ambiguity to assign intersubunit NOEs and determine the structure of a C_n symmetric homo-oligomer, given the subunit structure. By complete, we mean that our algorithm identifies structures representing to, within a user-defined similarity level, any structure consistent with available data (ambiguous or not). Here, by subunit structure we mean the ''bound'' structure of a monomer within the complex. The disambiguation of intra- versus intersubunit ambiguity makes it possible to determine the bound subunit structure prior to computing the oligomeric assembly (Oxenoid and Chou 2005).

Subunit ambiguity is a fundamental problem in dealing with homo-oligomers. While in some special cases it might be possible to manually resolve the ambiguity by identifying a self-consistent direction (e.g., in coiled coils; Wiltscheck et al. 1997), in general an NOE could involve a subunit i and any one of subunits $i + 1, i + 2, \ldots$ We call a determination of the relative positions of the subunits the ''subunit assignment'' of the NOE. For an n-mer with r NOEs, there are $(n - 1)^r$ possible combinations of subunit assignments, ruling out a naïve generate-and-test approach. Most previous attempts at resolving subunit ambiguity have done so simultaneously with resolving atom ambiguity, and have employed heuristic techniques (see below). As we discuss below, these approaches risk being trapped in local minima and missing correct assignments. A docking-based approach that sampled conformations could miss the native conformation and would require explicit enumeration of all combinations of assignments. One previous approach that rigorously deals with subunit ambiguity is AMBIPACK (Wang et al. 1998), which uses a branch-and-bound algorithm to compute rigid body transformations satisfying potentially ambiguous intersubunit distance restraints. AMBIPACK explicitly enumerates combinations of subunit assignments, but only for a limited number of the least ambiguous NOEs. However, unlike our approach, AMBIPACK does not take advantage of the ''closed-ring'' kinematics of symmetric homo-oligomers. Consequently, AMBIPACK uses randomized numerical techniques, and hence fails to be complete in identifying consistent structures, whereas our approach is complete in characterizing the entire space of consistent structures. Furthermore, AMBIPACK handles only a special case of subunit ambiguity, namely (counter)clockwise subunit ambiguity. By (counter)clockwise subunit ambiguity we mean that an NOE could either go clockwise or counterclockwise in a ring of monomers, resulting in two possible assignments. In the general case, arbitrary subunit ambiguity means that an NOE could go to any of the other $n - 1$ monomers, resulting in $n - 1$ possible assignments. Also, the presented AMBIPACK algorithm and results do not consider atom-ambiguous or spurious NOEs. On the other hand, our approach is able to resolve both arbitrary subunit ambiguity and atom ambiguity. We also show that our approach is able to handle spurious NOEs.

Chemical shift degeneracy in NMR spectra results in ambiguity as to which atoms are interacting in an NOE. We call a determination of the atoms the ''atom assignment'' of the NOE. A number of techniques, such as NOAH/DIAMOND (Xu et al. 1999), AUTOSTRUCTURE (Huang et al. 2003), and PASD (Juszewski et al. 2004), have been developed to address atom ambiguity in the context of monomer structure determination. These techniques follow an iterative assignment strategy: an initial set of unambiguous NOEs is used to generate some number of structures, which are then used to evaluate the consistency of atom assignments of the remaining ambiguous NOEs. The NOEs that are inconsistent with the ensemble are pruned. The remaining NOEs are incorporated in subsequent rounds of structure ensemble determination, and the process is repeated until no further improvements in atom assignments or structures can be obtained. An alternative approach to assign NOE restraints in monomer structure determination using a rotamer ensemble-based algorithm and residual dipolar couplings was developed by Wang and Donald (2005).

Previous approaches for dealing with both subunit and atom ambiguity in symmetric homo-oligomers perform a tight loop of NOE assignment and structure determination as described above for monomers. (Henceforth, when we refer to NOE assignment we mean intersubunit NOE

assignment.) Generation of structures consistent with a subset of NOEs is done by heuristic techniques such as simulated annealing, with additional constraints to ensure symmetry, e.g., by penalizing differences between the subunits (Revington et al. 2005) or by fixing multiple copies of the backbone and allowing only the side chains to move to fit the restraints (Oxenoid and Chou 2005). DYANA (Güntert et al. 1997) and CYANA (Herrmann et al. 2002; Güntert 2004) support this approach for homodimer structure determination, while the symmetry-ADR method in ARIA (Nilges 1993; O'Donoghue et al. 1993, 2000; Nilges et al. 1997) implemented in CNS/XPLOR (Brünger 1993; Brünger et al. 1998) is applicable to higher order oligomers.

The iterative techniques discussed above all follow a best-first strategy without backtracking, and thus must contend with the problem of multiple local minima. The large rearrangements potentially required to escape local minima are difficult to identify by local search techniques. The order in which assignments are chosen may affect the outcome, because the determined ensemble attempts to satisfy the chosen assignments, and the ensemble is then used to evaluate subsequent assignments. For example, CYANA uses the heuristic of choosing the most consistent NOEs at each iteration, but of course, this does not guarantee correctness. To ensure settling into the proper minimum, these techniques typically require a significant amount of over-restraint (e.g., ARIA requires >5 NOEs per residue). However, with higher order oligomers, available intersubunit data can be quite sparse, due to the size of the proteins and the decrease in sensitivity of peaks, along with the decrease in peak intensity due to double filtering and editing strategies. Finally, as the number of ambiguous NOEs increases, the ruggedness of the potential landscape increases, making it much more difficult and time-consuming to find valid solutions. Consequently, the level of atom ambiguity that current techniques can handle is also limited (e.g., CYANA requires a ^{1}H chemical shift match tolerance $\leq$0.03 ppm). Furthermore, the presence of subunit ambiguity in higher oligomers makes all NOEs inherently ambiguous and exacerbates the landscape ruggedness.

The ambiguity resolution algorithm developed here is an extension of our complete, data-driven search algorithm for C_n symmetric homo-oligomeric structure determination (Potluri et al. 2006). The key insight of that paper is that, given the structure of a subunit, the position and orientation of the symmetry axis determines the structure of a homo-oligomer with C_n symmetry: rotate the subunit n times by the angle of symmetry ($360°/n$) around the symmetry axis. Each possible conformation of the symmetric homo-oligomer is represented by a point in the four-dimensional *symmetry configuration space (SCS)*, $S^2 \times \mathbb{R}^2$ (orientation cross-position). The earlier paper then presents a branch-and-bound structure determination algorithm that identifies a set of 4D cells (hypercuboid regions in SCS), referred to as "consistent regions," ultimately identifying all conformations consistent with the data. However, our previous algorithm fails to handle atom ambiguity and arbitrary subunit ambiguity (subunit ambiguity involving nonadjacent subunits). As a result, it cannot use the additional constraint provided by all the ambiguous NOEs. The present paper significantly extends our earlier work with an algorithm that resolves atom and arbitrary subunit ambiguities. By exploiting the information in ambiguous NOEs, our new algorithm reduces the average variance in structures (over our previous approach) by a factor of 1.5 in our first test case (MinE) and a factor of 3.6 in our second test case (CCMP). These improvements make our approach applicable to the structure determination of a broad range of homo-oligomeric complexes.

In contrast with existing techniques, our approach is complete. (Recall that by complete we mean that it identifies structures representing, to within a user-defined similarity level, τ_0 Å, any structure consistent with the available data). We guarantee that the native conformations are always within τ_0 Å to at least one of the representative structures. Furthermore, the results of our algorithm are deterministic and not affected by the order in which NOEs are considered. In addition, while most approaches try to find one assignment for each ambiguous NOE restraint, our algorithm recognizes when additional ambiguity resolution will not improve structural precision and identifies the diminishing information content in NOEs. It might provably be the case that some NOEs do not need to be assigned. On the other hand, our approach can also recognize when different assignments lead to different structural classes, a possibility ignored by all other approaches. We also show that our approach can handle a reasonable number of spurious NOEs. In short, our ambiguity resolution algorithm extracts the information available from ambiguous NOEs, without bias in the possible assignments. Yet, it is still efficient, requiring time only linear in the number of ambiguous NOEs, the number of possible assignments for each, and the set of 4D cells in the SCS. Note that our approach focuses on intersubunit NOE assignment and currently cannot be applied to intrasubunit NOE assignment for monomer structure determination.

Results

We first summarize our computational results, with a problem formulation that enables identification of restraints that are inconsistent or redundant, and with a brief description of our algorithm. We then provide ambiguity resolution and structure determination results for two proteins, homo-dimeric MinE and homo-trimeric

CCMP, for which intersubunit NOE distance restraints and assigned chemical shift data were available in the BioMagResBank (Seavey et al. 1991; http://www.bmrb.wisc.edu). In both cases, inter- versus intrasubunit ambiguity had been resolved by isotopic labeling strategies. Thus, we use the subunit structure and focus on determining the complex from the remaining ambiguous intersubunit NOEs. The final two sections study, in the context of MinE, the effect of varying amounts of ambiguity as input to our approach, as well as the effect of spurious NOEs.

Problem formulation and approach

The problem that we must solve is the following. We are given a set of intersubunit NOEs (each possibly atom- and subunit-ambiguous), the subunit structure and the oligomeric number. Our goal is to determine a mutually consistent set of NOE assignments and SCS regions, such that the regions represent all conformations consistent with the assignments, and the assignments are all those consistent with the regions. Recall from the introduction that the SCS (symmetry configuration space) is the space of all symmetry axis parameters, each defining a homo-oligomeric complex structure. Thus, we seek to simultaneously assign NOEs and determine structure (by identifying regions in SCS), in a manner that guarantees that we find all consistent assignments and structures. In order to make such a guarantee, we must carefully formulate what it means for an NOE, or one of its possible assignments, to be *inconsistent* with respect to a given set of NOEs and the structures that they define. In the process, we are also able to formulate what it means to be *redundant*. In addition to helping us prove (see Materials and Methods) that our approach is correct, these formulations allow us to characterize (in the rest of the Results) the information content provided by a set of NOEs regarding a homo-oligomeric structure.

Let R be a set of (possibly ambiguous) NOEs. Let r_{kl} represent the l^{th} possible assignment of k^{th} NOE of R, and let $s_{kl} \subset S^2 \times \mathbb{R}^2$ be the region of the SCS in which r_{kl} is satisfied. (Recall that an assignment specifies the subunit and the atom involved in the NOE.) A determined structure must satisfy all ambiguous NOEs; it satisfies each by satisfying one or more of its possible assignments. In the SCS, this translates into finding the region $p(R)$ defined as:

$$p(R) = (s_{11} \cup s_{12} \cup s_{13} \cup \ldots) \cap (s_{21} \cup s_{22} \cup \ldots) \cap \ldots = \bigcap_{k=1}^{\|R\|} \left(\bigcup_{l=1}^{t_k} s_{kl} \right) \quad (1)$$

where t_k is the number of possible assignments for the k^{th} NOE.

Using Equation 1, we can give the definition for redundant and inconsistent NOEs and assignments. Let $Q \subset R$ be a set of NOEs. Let r_k be an (arbitrary) ambiguous NOE (not in Q) with t_k assignments. As above, let r_{kl} be its l^{th} possible assignment and s_{kl} be the region of the SCS in which the distance restraint r_{kl} is satisfied. We say r_{kl} is redundant with respect to Q if and only if $p(Q) \subset s_{kl}$. We say r_{kl} is inconsistent with respect to Q if and only if $p(Q) \cap s_{kl} = \varnothing$. We can extend these notions from assignments to NOEs as follows. Let $S_k = \cup_{l1}^{t_k} s_{kl}$ be the region of the SCS which satisfies at least one assignment of r_k (the k^{th} NOE). We then declare r_k to be redundant with respect to Q if and only if $p(Q) \subset S_k$, and declare r_k to be inconsistent with respect to Q if and only if $p(Q) \cap S_k = \varnothing$. Our algorithm eliminates inconsistent NOEs and assignments, and is able to detect redundant ones.

Our approach to identifying the mutually consistent, complete set of NOE assignments and SCS regions is summarized in Figure 1 and described in detail in the Materials and Methods. The following is a brief description of our approach. Our algorithm has two phases. In the first phase, branch-and-bound, we perform a search of the SCS using a subset of the NOEs (we show below that those with no atom ambiguity, if available, are the best choice). This phase identifies the *consistent regions* (a subset of the SCS), which contain symmetry axes defining all complexes consistent with the chosen subset of the NOEs. In the second phase, ambiguity resolution, we sequentially consider ambiguous NOEs and their possible assignments, using an ordering criterion that chooses informative NOEs. This phase identifies the *ambiguity-consistent regions* (ACR), which contain symmetry axes defining all complexes consistent with some assignment for each NOE. As we formally defined above, this process allows us to label assignments as inconsistent (violated everywhere in the ambiguity-consistent regions) or redundant (does not provide any additional constraint for the regions).

We represent both consistent regions and ambiguity-consistent regions with cells in a hierarchical decomposition of the symmetry configuration space. Thus, the volume of these cells allows us to characterize the progress in structure determination, as the algorithm focuses on consistent symmetry axes. We generate representative conformations from these cells such that any structure in any cell is within τ_0 Å RMSD (the user-defined similarity level) of at least one representative structure. We refer to those representative structures that have good van der Waals packing as "*well-packed satisfying*" (WPS) structures. The average variance in atomic positions in these conformations allows us to evaluate the structural precision attained by the two phases of the algorithm.

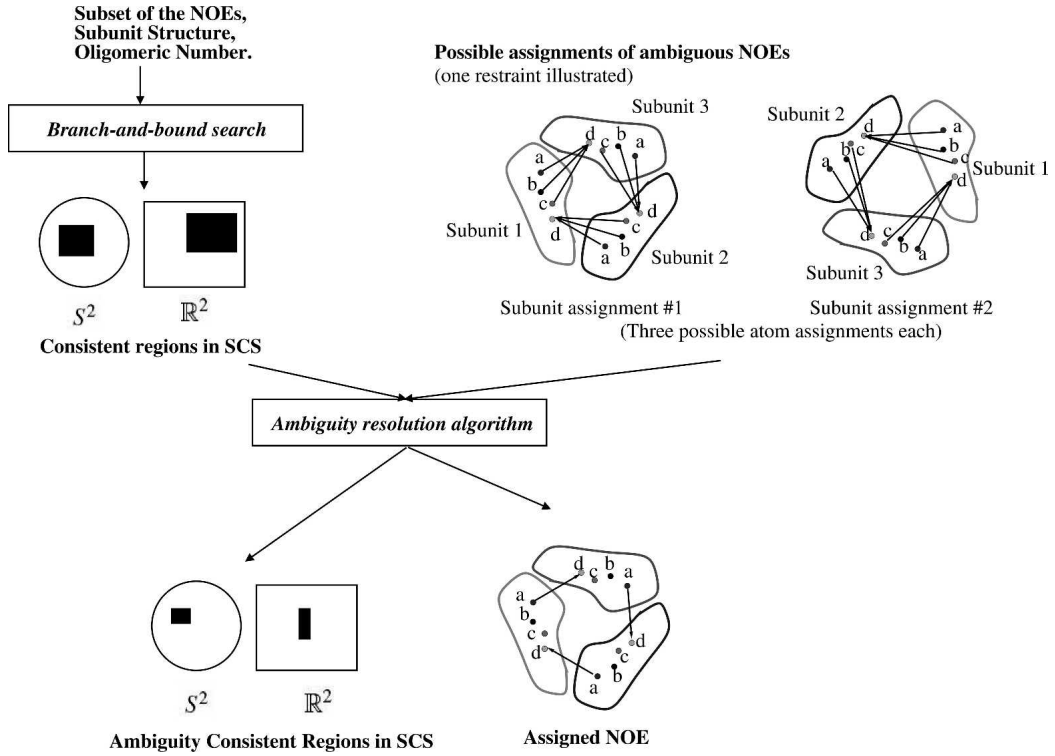

Figure 1. Resolving ambiguous NOEs in structure determination of symmetric homo-oligomers. Different possible assignments (arising from subunit and atom ambiguity) for one ambiguous NOE are illustrated for a trimer. The NOE could be ordered between subunits as 1–2–3 or 1–3–2, and has chemical shift degeneracy between atoms a, b, and c. (Phase 1) Consistent regions in the symmetry configuration space (SCS), the space of symmetry axis parameters, are obtained using our previously developed branch-and-bound algorithm (Potluri et al. 2006), taking a subset of available NOEs (typically atom-unambiguous NOEs if any) as input. The four-dimensional SCS is illustrated as two 2D regions, a sphere representing the orientation space S^2, and a square representing the translation space $\mathbb{R}^2$. (Phase 2) our ambiguity resolution algorithm takes as input the consistent regions output from our branch-and-bound algorithm and possible assignments of the ambiguous NOEs to determine a mutually consistent, complete set of NOE assignments and ambiguity consistent regions (ACR) in SCS.

Ambiguity resolution and structure determination of homo-dimeric MinE

The homo-dimeric topological specificity domain of *Escherichia coli* MinE (King et al. 2000) has a novel dimeric αβ-sandwich fold that has previously been determined (PDB ID 1EV0) by using both DYANA (Güntert et al. 1997) and the NCS-symmetry potential of XPLOR (with the ARIA method) (Brünger et al. 1998). Each subunit has 50 residues, and a total of 183 intersubunit NOE restraints were deposited. We use as the subunit the first chain of the reference structure (stated by the authors to be the best representative conformer). Subunit ambiguity is absent in homo-dimers. Atom ambiguity was simulated for each of the intersubunit NOE restraints, according to the deposited chemical shifts, using a chemical shift match tolerance for ^{1}H of $\varepsilon = 0.04$ ppm. This resulted in atom ambiguity for 168 of the 183 NOEs, as illustrated in Figure 2A. Note that there is ambiguity as high as 12 (i.e., 12 possible assignments) for two of the NOEs.

We first ran our branch-and-bound algorithm for a complete SCS search with the 15 atom-unambiguous NOEs. The consistent regions output from this step included a volume of 0.030 Å^2 rad^2 represented by 70 cells in SCS. We then ran our ambiguity resolution algorithm. Figure 3 shows the decrease in the SCS volume over the number of NOEs chosen. The running time for the previously reported branch-and-bound search (Potluri et al. 2006) on a Intel Pentium Linux desktop machine with a 3.20-GHz CPU was <9 h, while that for the ambiguity resolution algorithm was <1 h. The SCS volume after 30 NOEs were chosen using our ordering criterion was 0.025 Å^2 rad^2. No further decrease in SCS volume occurred after this, indicating that the NOEs chosen later are redundant with respect to the chosen ones. Note that ~90% of the reduction in volume occurred after three NOEs were chosen, an indication that our ordering criterion does focus on more informative NOEs. Our ambiguity resolution algorithm eliminates the inconsistent assignments and Figure 2B shows

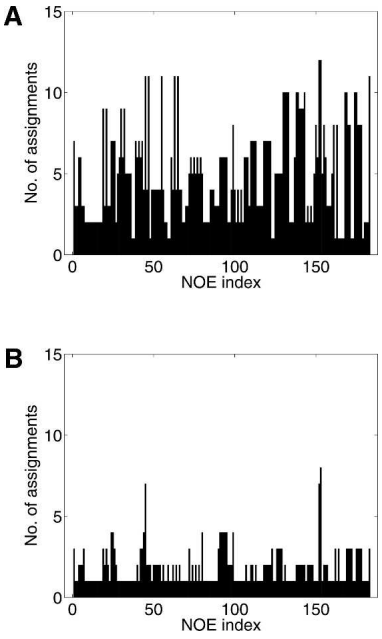

Figure 2. Ambiguity resolution for homo-dimeric MinE (1EV0). Plotted is the number of assignments for each NOE after the branch-and-bound algorithm but before ambiguity resolution (*A*) and after ambiguity resolution (*B*).

the number of possible assignments remaining for each NOE after all the inconsistent assignments have been eliminated. On comparison with Figure 2A, we see that the ambiguity is considerably reduced, from an average of 5.0 assignments per NOE to an average of 1.9.

Figure 4 illustrates the effect of ambiguity resolution on the resulting conformations (WPS structures): The average variance is decreased from 0.34 $Å^2$ (WPS structures after branch-and-bound but before ambiguity resolution) to 0.22 $Å^2$ (after ambiguity resolution), and the average RMSD to the reference structure is reduced from 0.95 Å to 0.73 Å. There are no experimental NOEs observed between atoms in the N and C termini; consequently, the termini display higher uncertainty. Note that ambiguity resolution does decrease the uncertainty in these regions, by way of restraint from NOEs involving other atoms. While the assignment is not unique (Fig. 2B shows NOEs with as many as eight possible assignments remaining), the remaining assignments are consistent with the determined set of WPS structures and their effect is minimal.

Our algorithm employs a particular criterion (described in Materials and Methods) to determine the order in which to consider NOEs. We show in Materials and Methods (Claim 1) that the order does not affect the results, but only the efficiency in obtaining them. To evaluate the efficiency of our ordering criterion, we compared it against a strategy in which the order of NOEs is chosen randomly. Figure 3 shows that our criterion prunes the SCS much faster than random ordering. Our ordering criterion has fully constrained the SCS volume with the first 30 NOEs, earlier than all but one of 100 random runs. We show in Materials and Methods (Claim 2) that the time spent for each NOE is linear in the number of cells tested for consistency. Our ordering criterion prunes inconsistent cells much earlier than random ordering, thereby allowing us to spend time on the regions most likely to be part of the ACR. Figure 5 shows a histogram of the total number of consistency tests for the 100 random orderings. The number of consistency tests using our ambiguity resolution algorithm is also shown. The figure illustrates that our ordering criterion requires many fewer consistency tests than most of the 100 random trails, and is thus much faster.

Ambiguity resolution and structure determination of homo-trimeric CCMP

The second test case, the trimeric coiled-coil domain of chicken cartilage matrix protein (CCMP) (Wiltscheck et al. 1997), displays subunit ambiguity in addition to atom ambiguity. Each subunit has 47 residues, and a total of 49 intersubunit NOE restraints were deposited. The structure was originally determined using XPLOR (Brünger et al. 1998) (PDB ID 1AQ5), with subunit ambiguity resolved by manual identification of a self-consistent subunit identity for all 49 intersubunit restraints. We used as the subunit the first chain of the reference structure (stated by the authors to be the best representative conformer). We ignored the subunit assignment of the authors (Wiltscheck et al. 1997), and simulated atom ambiguity as in the homo-dimer case (with $\varepsilon = 0.04$ ppm). Figure 6A illustrates the number of assignments that arise from subunit and atom ambiguity for each of the 49 NOEs. Two of the NOEs have as many as 36 possible assignments, when considering the combination of subunit and atom ambiguity.

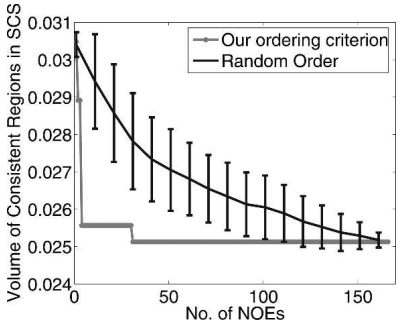

Figure 3. Progress in ambiguity resolution for homo-dimeric MinE (1EV0). The NOE considered at each step is chosen either by our ordering criterion (gray line with hexagons) or at random (solid black line; mean and error bars over 100 random trials). Plotted is the decrease in total volume of the SCS cells with number of NOEs chosen.

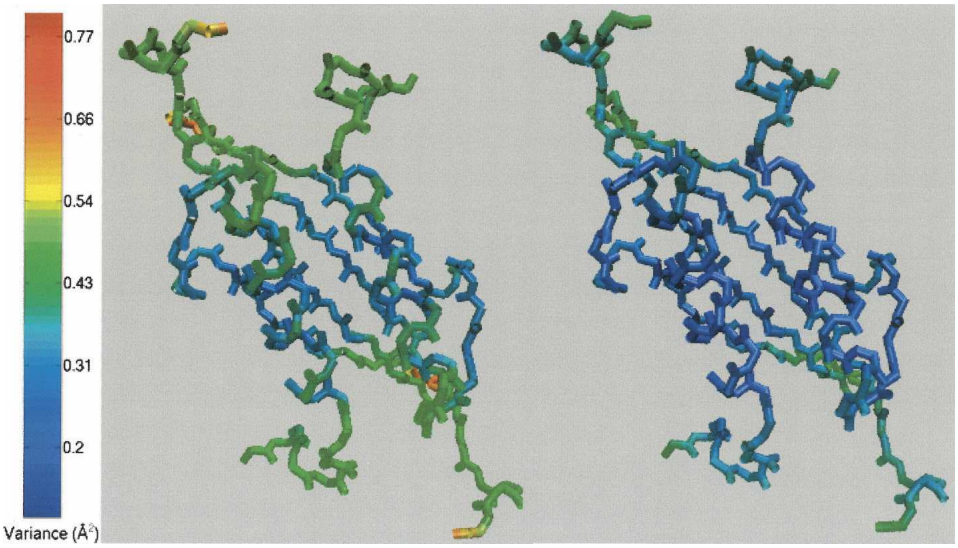

Figure 4. Variance in WPS structures of homo-dimeric MinE (1EV0): (*Left*) after the branch-and-bound algorithm but before ambiguity resolution and (*right*) after ambiguity resolution. Each backbone atom is colored by the variance in the position of the atom according to the shown color scale.

A complete SCS search was performed with the five NOEs that had no atom ambiguity. The search identified a volume of 0.27 $Å^2$ rad^2 represented by 4006 cells in SCS. Note that the presence of subunit ambiguity caused a significantly larger set of cells. We then ran our ambiguity resolution algorithm. Our ambiguity resolution algorithm reduced the volume to 0.037 $Å^2$ rad^2. The running time for the branch-and-bound search on a Intel Pentium Linux desktop machine with a 3.20-GHz CPU was around a couple of days, while that for the ambiguity resolution algorithm was around 7 h. As Figure 7 shows, no further decrease in SCS volume occurred after 32 NOEs were chosen, with the bulk of the reduction (99%) occurring after five NOEs were chosen. The flat regions in Figure 7 for number of chosen NOEs 5–10 and 14–28 indicate that the NOEs considered then were redundant in the presence of the already handled NOEs. Our ambiguity resolution algorithm eliminated all the inconsistent assignments and Figure 6B illustrates the surviving assignments, significantly reduced (from average of 18.2 to average of 4.3) from Figure 6A.

Figure 8 illustrates the impact of ambiguity resolution on WPS structures. Even though many NOEs have not been resolved to a single assignment, the final set of WPS structures has high precision: the average variance is decreased from 1.92 $Å^2$ (after the branch-and-bound but before ambiguity resolution) to 0.54 $Å^2$ (after ambiguity resolution), and the average backbone RMSD to the reference structure is reduced from 1.82 Å to 0.6 Å. As Figure 8B shows, our ambiguity resolution algorithm reduced the uncertainty in the positions of all the atoms to <1 $Å^2$. Our algorithm reduced the high uncertainty that was present at the N and C termini of the chains after the branch-and-bound algorithm to <1 $Å^2$.

Figure 7 compares the number of consistency tests under our ordering criterion versus under a random choice, and Figure 9 shows the distribution of number of cells checked for consistency. As with the homo-dimer, the ordering criterion leads to significantly less work than most of the random trails.

Effect of maximum ambiguity level in branch-and-bound input

The results so far have used only the atom-unambiguous NOEs as input to the branch-and-bound phase of our algorithm. Here we study the use of NOEs with varying

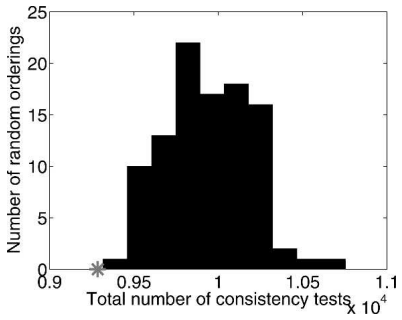

Figure 5. Histogram over 100 random trials of the computational cost of ambiguity resolution for homo-dimeric MinE (1EV0). The asterisk indicates the number of consistency tests made using our ordering criterion.

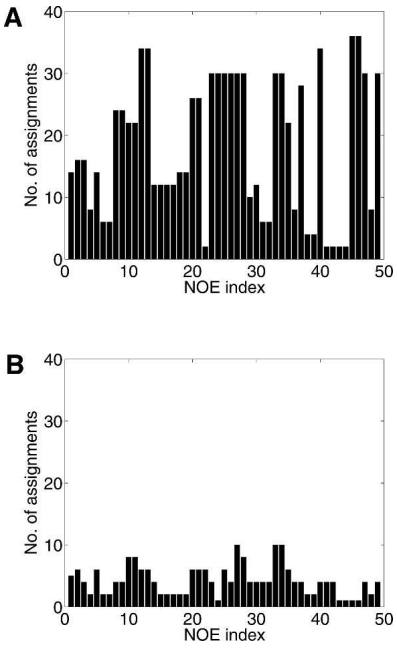

Figure 6. Ambiguity resolution for homo-trimeric CCMP (1AQ5). Plotted is the number of assignments for each NOE after the branch-and-bound algorithm but before ambiguity resolution (*A*) and after ambiguity resolution (*B*).

amounts of ambiguity for that phase. Note that using any different set of NOEs as input to the branch-and-bound phase will always ultimately lead to the same results. It affects efficiency, not correctness. We define the set of NOEs with maximum ambiguity level k to include all NOEs having k or fewer possible assignments. We would expect that, with increasing maximum ambiguity level, more work would be done in the branch-and-bound search (since it has to deal with more ambiguity) and less work would be done in the ambiguity resolution algorithm (since the branch-and-bound search has eliminated some ambiguity).

Figure 10 shows that this is indeed the case for MinE, measuring the amount of work as the number of consistency tests performed. The figure also shows that the extra work in the branch-and-bound search with increased maximum ambiguity level is not worth it—the number of tests summed across both phases increases with the maximum ambiguity level. When all NOEs are used as input to the branch-and-bound search, the total number of consistency tests required is 7.8 times more than the number required when only atom-unambiguous NOEs are used. It is more efficient to start with only the atom-unambiguous NOEs.

Our approach is applicable even when no atom-unambiguous NOEs are available. To demonstrate this, we ignored the 15 atom-unambiguous NOEs for MinE, and ran the branch-and-bound search with the 34 (out of 168) NOEs with maximum ambiguity level 2. The output from this step was an SCS volume of 0.042 $Å^2$ rad^2 represented by 94 cells. We then ran our ambiguity resolution algorithm with the remaining 134 NOEs. This reduced the SCS volume to 0.034 $Å^2$ rad^2. The ambiguity in the 168 NOEs was reduced from an average of 5.3 assignments per NOE to an average of 2.0. The average variance in WPS structures was decreased from 0.37 $Å^2$ (after branch-and-bound but before ambiguity resolution) to 0.22 $Å^2$ (after ambiguity resolution). The final structural precision after the ambiguity resolution algorithm, 0.22 $Å^2$ is the same as when we started with the 15 atom-unambiguous NOEs. Thus, there was enough redundant information in the atom-ambiguous restraints to compensate for the loss of the 15 atom-unambiguous ones.

Effect of spurious NOEs

So far, we have assumed that each NOE peak is true; that is, that every NOE has at least one correct assignment. However, our approach is readily extended to handle a small number of spurious NOEs, NOEs for which all assignments are wrong. In this case, we eliminate a cell when at least a specified number (which is one in the presented algorithm) of the NOEs are violated. An NOE is identified as spurious if it is inconsistent with all the remaining cells of the ACR. Note that we may not be able to identify those spurious NOEs that are consistent with the remaining cells of the ACR.

We tested this approach with MinE, simulating spurious NOEs from the reference structure by randomly choosing pairs of protons that were >10 Å apart. Atom ambiguity was simulated for each of the spurious NOE as before (with ^{1}H chemical shift match tolerance $\varepsilon = 0.04$ ppm). We performed 100 random runs for each number of spurious NOEs, from one to five.

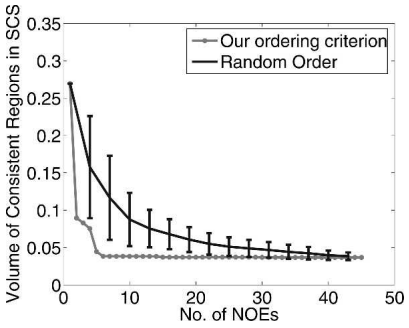

Figure 7. Progress in ambiguity resolution for homo-trimeric CCMP (1AQ5). The NOE considered at each step is chosen either by our ordering criterion (hexagons) or at random (solid line; mean and error bars over 100 random trials). Plotted is the decrease in total volume of the SCS cells with number of NOEs.

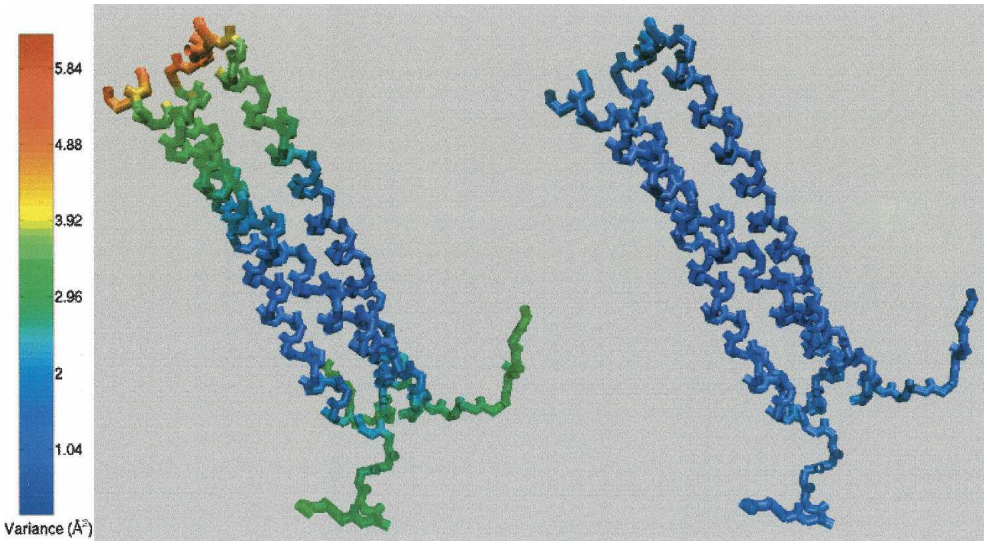

Figure 8. Variance in WPS structures of homo-trimeric CCMP (1AQ5): (*Left*) after the branch-and-bound algorithm but before ambiguity resolution and (*right*) after ambiguity resolution. Each backbone atom is colored by the variance in the position of the atom according to the shown color scale.

Figure 11A illustrates the detection rate for the spurious NOEs. In the case when one spurious NOE is added, the detection rate is 93%. This means that in 93% of the cases, the spurious NOE has no effect on the remaining cells in the ACR or the remaining assignments. The sets of mutually consistent cells and assignments are the same as that with no spurious NOEs. The failure of detection in the remaining seven cases is due to the presence of an atom assignment for the spurious NOE that is consistent with the remaining conformations. The detection rate decreases to 62% as the number of spurious NOEs is increased from one to five.

However, as illustrated in Figure 11B–D, even if a spurious NOE cannot be detected, it has minimal effect on the volume in SCS, the structural precision (indicated by average variance) and the structural quality (indicated by the average backbone RMSD to the reference structure). Note that even with as many as five spurious NOEs, the mean of the average variance after the ambiguity resolution algorithm is around 0.30 $Å^2$, and is less than the average variance of the ACR before use of the ambiguity resolution algorithm (0.34 $Å^2$). This indicates that we can obtain information and improve structural precision, in spite of the presence of spurious NOEs.

Conclusion

We have developed a complete algorithm that resolves both the atom and subunit ambiguity inherent in NOEs to assign NOE peaks and determine structures of symmetric homo-oligomers by NMR. Our algorithm determines a mutually consistent, complete set of NOE assignments and conformations representing (to within a user-defined similarity level) any structure consistent with the available data (ambiguous or not). Our approach is applicable to any C_n symmetric homo-oligomer, and considers all possible assignments for available NOEs. Using our approach, atom ambiguity was reduced by a factor of 2.6 in our first test case (MinE) and combined subunit/atom ambiguity by a factor of 4.2 in our second test case (CCMP). Incorrect NOE assignments tend to be mutually inconsistent, and therefore are eliminated by our algorithm. Our ambiguity resolution algorithm obtained structures with high precision: the average variance in positions of atoms is reduced to 0.22 $Å^2$ for MinE and 0.54 $Å^2$ for CCMP. Our sequential strategy chooses the NOE that appears informative and focuses on cells (hypercuboid regions in SCS) that potentially satisfy all

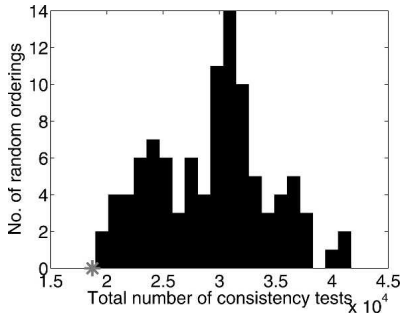

Figure 9. Histogram over 100 random trials of the computational cost of ambiguity resolution for homo-trimeric CCMP (1AQ5). The asterisk indicates the number of consistency tests made using our ordering criterion.

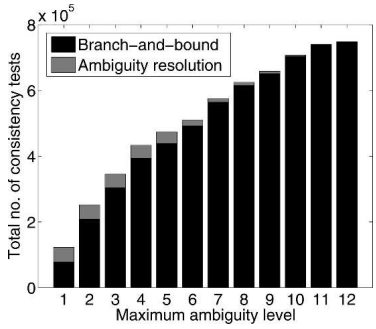

Figure 10. Effect of the amount of ambiguity in the branch-and-bound search, for MinE (1EVO). The maximum ambiguity level limits the NOEs used in the search phase. The plot shows the total number of consistency tests during the branch-and-bound search and during the ambiguity resolution algorithm.

the NOEs. This allows it to prune a significant number of bad cells (~90%) early, and hence, results in greatly reduced time complexity. In the two cases studied here, the branch-and-bound plus ambiguity resolution required only hours to days on a single processor in order to guarantee complete consideration of all possible solutions. Complete computation of all solutions avoids bias in the search, as well as any potential for becoming trapped in local minima, both of which are problems inherent in existing heuristic approaches. The run time can potentially be further reduced by distributing the computation over a commodity cluster with a straightforward extension of our implementation. Our sequential strategy also helps elucidate the information content in ambiguous NOEs; many NOEs may be redundant with other (previously handled) NOEs. We identify that 138 of the 183 NOEs in MinE and 12 of the 49 NOEs in CCMP are redundant with respect to the chosen NOEs and cause no further improvement in structural precision. We have also shown that our approach can work with NOEs at different maximum ambiguity levels provided as input to the branch-and-bound search, and can handle a reasonable number of spurious NOEs.

NOE assignment can be difficult to fully automate. At the same time, the structure determination of symmetric proteins by NMR can be challenging. It is interesting that by combining these two difficult problems, an algorithm can be obtained that solves both simultaneously, and that enjoys guarantees on its completeness and complexity. Our approach could be extended to other kinds of symmetry, such as a dimer of dimers, a trimer of dimers, or improper symmetry, by defining appropriate configuration spaces and searching them in an analogous manner. For instance, in the case of dimer of dimers (D_2 symmetry), the space of symmetry axes would be

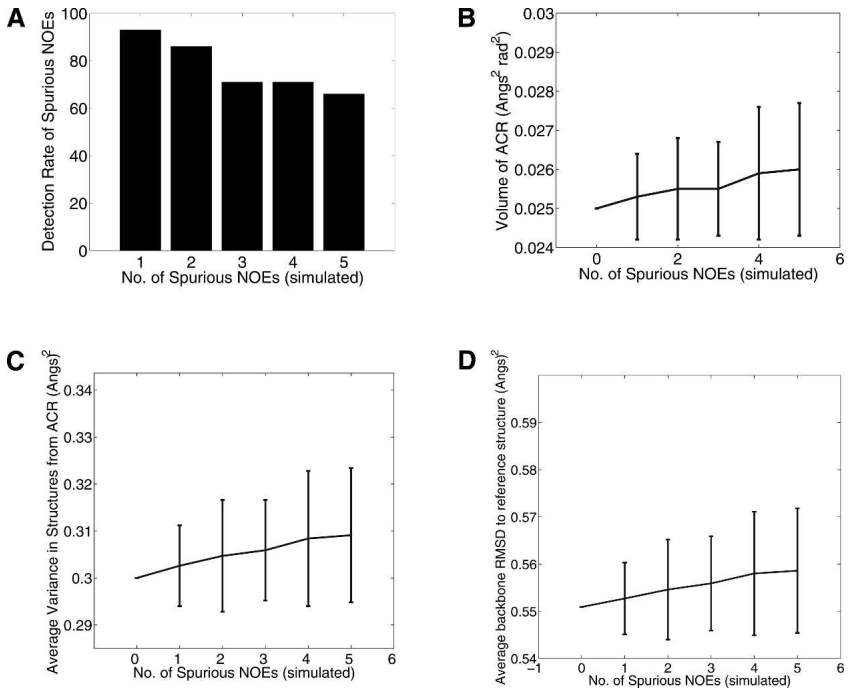

Figure 11. Effect of spurious NOEs on ambiguity resolution for MinE (1EVO). The effect was studied on an increasing number of simulated spurious NOEs, with 100 random runs each. (*A*) The detection rate of spurious NOEs. (*B*) The mean and error bars for the volume of ACR. (*C*) The mean and error bars for the average variance of representative structures of ACR. (*D*) The mean and error bars for the average backbone RMSD of representative structures of ACR to the reference structure.

represented by seven dimensions (four dimensions for one symmetry axis and three dimensions for the symmetry axis perpendicular to the first), and bounds analogous to the bounds we have previously obtained for the four-dimensional case (Potluri et al. 2006) would have to be developed. The developed bounds would then be used in the hierarchical subdivision of the configuration space and ambiguity resolution of the NOEs. We believe our algorithm could contribute to the structural characterization of symmetric membrane proteins and other large homo-oligomers by NMR.

Our software can be freely obtained for academic use by request from the corresponding authors.

Materials and methods

Representation of assignment

We first formalize the representation of subunit and atom assignments discussed in the introduction (Fig. 1).

Subunit assignment

An NOE could capture an interaction between any pair of protons in any pair of subunits. Under C_n symmetry, each interaction is "mirrored," i.e., holds between each pair of subunits that are indistinguishable under C_n rotation. Thus, a subunit assignment for an NOE specifies the difference (1 to $n-1$) in the positions of the involved subunits within the symmetric order.

Atom assignment

In typical approaches to NMR structure determination, the chemical shift assignments of the various atoms are obtained in an earlier analysis, and are then used to determine which atoms' interactions might have generated the NOE peaks. The possible identities of the interacting protons are determined by comparing the chemical shift coordinates of an NOE peak with the assigned chemical shifts (the same for each subunit, by C_n symmetry). In the 3D spectra used here, there are two coordinates (^{15}N/^{13}C and ^{1}H) by which to determine one side of the interaction and one coordinate (^{1}H) for the other. Since two coordinates generally result in a unique assignment, we here consider only one-sided atom ambiguity, although our approach can readily be generalized. Thus, an atom assignment for an NOE specifies the identity of the ambiguous atom, from the entire set A of atoms (here we consider only protons) in a monomer. Typically, a user-specified match tolerance ε (e.g., 0.04 ppm) limits the possible protons to those with similar chemical shift to the peak coordinate.

In summary, the assignment for an NOE specifies a pair $(i,j) \in Z_{n-1} \times A$, where i assigns the relative positions of the subunits in the complex and j assigns the proton in the subunit. Here, Z_{n-1} represents S_n, the symmetric group of n elements. Each element of S_n is represented by an integer in $\{1,\ldots, n-1\}$.

Ambiguity resolution algorithm

Although we would like to compute exactly the region $p(R)$ (Equation 1) defining all consistent structures, it is a semi-algebraic set characterized by high-degree polynomials that are expensive to solve exactly. Our ambiguity resolution algorithm computes a superset of $p(R)$ by using the conservative analytical bounds developed in Potluri et al. (2006) to test satisfaction of possible NOE assignments. We say that our method is complete because every element of $p(R)$ is provably guaranteed to be in the superset returned by our algorithm. More specifically, our method takes as input an initial region of the SCS (represented as a set of nonoverlapping cells, each of which is a hypercuboid region in $S^2 \times \mathbb{R}^2$, in a hierarchical decomposition), and outputs the intersection between that region and a superset of $p(R)$. In this way, it is possible to sequentially compute a superset of $p(R)$ by sequentially "intersecting" more and more (possibly ambiguous) NOEs.

We explain our algorithm by first describing the relationship between our previous work and the present algorithm (refer again to Fig. 1). Our previously developed branch-and-bound structure determination algorithm (Potluri et al. 2006) takes as input available intersubunit NOEs, the oligomeric number n, and the subunit structure (previously determined from the intra-subunit restraints, as is possible under the labeling strategies discussed in the introduction) and identifies a set of 4D cells, which we call the "consistent regions." The consistent regions contain all conformations consistent with the data. The algorithm assumes that the backbone displays exact C_n symmetry, but we allow for asymmetry and flexibility in the side chains when we subject the structures to energy-minimization (discussed below). The algorithm employs a conservative bound that eliminates a cell only if all conformations it represents are guaranteed to violate an NOE. The search algorithm recursively subdivides the cells, terminating when all the structures within a cell are within a user-defined similarity level, τ_0 Å of each other.

In this paper, we use that algorithm to determine the consistent regions for a subset (typically atom-unambiguous) NOEs. (We note that subunit ambiguity always exists for oligomers with more than two subunits.) We then apply our novel ambiguity resolution algorithm (Fig. 12) that considers each of the remaining ambiguous NOEs sequentially, that is, one after the other. At each step, we use our ordering criterion (next section) to choose the NOE considered most informative. Each SCS cell of the consistent regions is checked, using our conservative bound, to ensure that some assignment of the NOE is consistent with the structures represented by the cell. A cell is eliminated if every assignment of any of the NOEs is violated by all the conformations represented by that cell. We then continue with the next NOE chosen using our ordering criterion and the remaining cells. This process of choosing, updating, and resolving is continued until all the NOEs have been considered. The remaining cells form the ambiguity-consistent regions (ACR). Finally, we resolve the ambiguity in the NOEs by eliminating assignments inconsistent with the ACR. Our ambiguity resolution algorithm hence identifies every mutually consistent set of assignments and ambiguity-consistent regions.

After ambiguity resolution, we choose representative structures from the ACR such that every conformation in the ACR is within an RMSD of τ_0 Å to at least one representative structure. We evaluate the representatives for quality of NOE restraint satisfaction, and then energy-minimize them and evaluate them for quality of van der Waals packing, ultimately identifying the set of well-packed satisfying (WPS) structures.

NOE ordering criterion

We use a particular ordering criterion to choose the next NOE in our sequential strategy. Ordering of the NOEs only affects the efficiency of our approach, and not the result.

Input:
R : Set of NOEs
$U \subset R$: Subset of the NOEs (typically atom-unambiguous NOEs)
C : Set of cells from branch-and-bound with U
$N : R \rightarrow \mathbb{Z}_{n-1} \times A$: Possible assignments for each NOE
due to subunit ambiguity (subunit offset 1 to $n-1$)
and atom ambiguity (proton in the set A).

Output:
$C' \subset C$: Remaining cells
$N' \subset N$: Remaining assignments

Algorithm:
initialize $R' \leftarrow R - U,\ C' \leftarrow C$
while $R' \neq \emptyset$
// Determine the next NOE to assign according to our criterion
$S \leftarrow$ representative structures for C'
$q \leftarrow \arg\max_{r \in R'} \text{average}_{a \in N(r)} |\{s \in S \mid r \text{ under assignment } a \text{ is violated in } s\}|$ (*)
$R' \leftarrow R' - \{q\}$

// Remove inconsistent cells
$C' \leftarrow \{c \in C' \mid \exists a \in N(q) \text{ s.t. } c \text{ is consistent with } q \text{ under assignment } a\}$
end while

// Remove inconsistent assignments
$N' \leftarrow \{(r, a) \mid r \in R, a \in N(r), \exists c \in C' \text{ s.t. } c \text{ is consistent with } r \text{ under assignment } a\}$

return N', C'

Figure 12. Ambiguity resolution algorithm.

Claim 1

Reordering the NOEs in R does not affect the completeness of our sequential algorithm.

Proof sketch

This is a direct consequence of the fact that set intersection is commutative, and the fact that our method returns a superset of $p(R)$ (Equation 1). □

However, the order does impact the efficiency, and we want to identify the NOE that causes the largest decrease in the SCS, in order that fewer consistency tests are performed. Identifying such an NOE will decrease the number of satisfaction checks required for the remaining assignments, and hence decreases the overall time complexity. To find the next NOE to assign, we first generate a set of representative structures, one from each cell. For each assignment of an NOE, we compute its scores as the number of representative structures that violate the NOE under the assignment. This is indicative of the portion of the SCS that would be eliminated under the assignment. Each NOE's score is then the average of the scores of all of its possible assignments (the line marked (*) in Fig. 12). We choose the next NOE to assign as the one with the maximum score. Although our ordering criterion does not guarantee optimal efficiency, our results show that it does very well in practice, and allows for the identification of redundant NOEs. We also tried evaluating an NOE by the minimum of its assignment scores, but found this approach to be inferior (data not shown).

Computational complexity

When there are r NOEs with a assignments each, we have a^r possible combinations of assignments. An advantage of our algorithm is that it avoids the potentially exponential enumeration of all a^r possible combinations of NOE assignments.

We first run the branch-and-bound search with a subset of the available NOEs (typically atom-unambiguous NOEs) as input. The output from the search is then input to our ambiguity resolution algorithm to ultimately return a mutually consistent set of ACR and NOE assignments.

Claim 2

Given c cells as consistent regions from the branch-and-bound search with a subset of the NOEs, the time complexity of our algorithm is $O(rac)$, where r is the number of ambiguous NOEs, and a is the number of assignments for each NOE.

Proof sketch

Let t be the time complexity for checking satisfaction of an NOE in one cell. Assigning a particular NOE requires a satisfaction test in each of the cells for each of the assignments (i.e., whether some conformation in the cell is consistent with the assigned NOE) and thus a total time of $O(atc)$ in the worst case where no assignments and no cells have been eliminated. The time complexity for a satisfaction test is constant, $t = O(1)$ (see Appendix for proof). Therefore, assigning one NOE has a time complexity of $O(ac)$. Our algorithm considers a total of r NOEs, and in the worst case each is independent of the others (i.e., no pruning occurs), for a total time of $O(rac)$. □

The number of cells, c, depends on the complexity of the well-packed satisfying regions in the SCS. These regions are bounded by algebraic hypersurfaces. In principle, a worst-case bound could be obtained based on the combinatorial complexity of this four-dimensional arrangement (Edelsbrunner 1987). In practice, we have characterized c empirically for the test cases in this paper ($c = 70$ for the homo-dimeric MinE and $c = 4006$ for the homo-trimeric CCMP with atom-unambiguous NOEs as input) and report that our algorithm is efficient in practice. In general, for ambiguous NOE assignment problems, the bottleneck in combinatorial complexity lies in the potentially exponential number a^r of possible assignments, and for this reason our methodology has focused on developing an algorithm that can provably make correct and consistent assignments while only considering a linear number of possible assignments, ar.

Acknowledgments

We acknowledge members of the C.B.K. lab for helpful discussions, and Serkan Apaydin from B.R.D. lab for comments on the manuscript. This work was supported in part by the following grants to B.R.D., National Institutes of Health (R01 GM 65982) and National Science Foundation (EIA-9802068 and EIA-0305444); and to C.B.K, National Science Foundation (IIS-0444544 and IIS-0502801).

Appendix

Theorem 1

The time complexity for checking satisfaction of a possible NOE assignment in a cell of the 4D SCS is $O(1)$.

Proof

Consider checking satisfaction of a restraint of the form $\|\mathbf{p} - \mathbf{q}'\| \leq d$ ($\mathbf{p}$ and $\mathbf{q}'$ are atoms on different subunits and d is

the distance of the restraint) in a cell of the 4D SCS, $G = G_s \times G_r \subset S^2 \times \mathbb{R}^2$. This requires determining if there is a nonempty intersection between a ball (centered at **p** and of radius d) and $G\mathbf{q}$, the region of possible positions of $\mathbf{q}'$ when the symmetry axis is in G. Since $G\mathbf{q}$ is hard to compute exactly, we approximate it by computing a bound on $G_s\mathbf{q}$ at the four corners of G_r and then taking the convex hull of the four bounds (Section 2.1.1 of Potluri et al. 2006). The resulting convex hull is a superset of $G\mathbf{q}$, which we use in our conservative tests for restraint satisfaction. The bound on $G_s\mathbf{q}$ is obtained as an axis-aligned box (AABB). The time complexity, t, for checking satisfaction of a restraint in a cell, would then be the sum of (a) the time complexity, t_b, for computing the bound on $G_s\mathbf{q}$ at each of the four corners of G_r, (b) the time complexity, t_c, for computing the convex hull of the bounds at the corners, and (c) the time complexity, t_i, for computing the intersection between the convex hull and a ball.

(a) The time complexity, t_b, for computing a bound on $G_s\mathbf{q}$ includes evaluating an analytical expression to compute the center and radius of a bounding ball (Equations 1 and 2 in Section 2.1.1 of Potluri et al. 2006) and computing an axis-aligned box (AABB) of the intersection of the bounding ball and a sphere. This requires constant time, $t_b = O(1)$. Since the bound on $G_s\mathbf{q}$ must be obtained at each of the four corners of G_r, $t_b = 4 * O(1) = O(1)$.

(b) The time complexity, t_c, for computing the convex hull of the AABBs obtained at the four corners of G_r would be $O(s \log s)$ in the worst case, where s is the number of input points to compute the hull. Since we are finding the convex hull of four AABBs, each with eight vertices, s in our case would be 32. Hence, $t_c = O(1)$.

(c) The time complexity, t_i, for computing the intersection between the convex hull and a ball is proportional to the number of triangles of the convex hull. For each triangle, we test if it intersects the ball. Since the complexity for testing each triangle is constant, $t_i = O(f)$, where f is the number of triangles of the convex hull. Using Euler's formula, we can prove that the number of triangles for the convex hull in 3D is at most $2p - 4$ (de Berg et al. 2000). Here, p is the number of vertices of the convex hull, which is at most the number of input points, 32 in our case. Hence, $f \leq 2 * 32 - 1$, obtaining $t_i = O(1)$.

Hence, the time complexity for checking satisfaction of a possible NOE assignment in a cell of the 4D SCS, $t = t_b + t_c + t_i = O(1)$.

References

Brünger, A.T. 1993. *XPLOR: A system for X-ray crystallography and NMR*. Yale University Press, New Haven, CT.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. 1998. Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54:** 905–921.

de Berg, M., van Krevald, M., Overmars, M., and Schwarzkopf, O. 2000. *Computational geometry*. Springer-Verlag, Berlin.

Edelsbrunner, H. 1987. *Algorithms in combinatorial geometry*. Springer-Verlag, Berlin.

Fiorito, F., Hiller, S., Wider, G., and Wüthrich, K. 2006. Automated resonance assignment of proteins: 6D APSY-NMR. *J. Biomol. NMR* **35:** 27–37.

Güntert, P. 2004. Automated NMR protein structure calculation with CYANA. *Methods Mol. Biol.* **278:** 353–378.

Güntert, P., Mumenthaler, C., and Wüthrich, K. 1997. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273:** 283–298.

Herrmann, T., Güntert, P., and Wüthrich, K. 2002. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319:** 209–227.

Huang, Y.J., Swapna, G.V., Rajan, P.K., Ke, H., Xia, B., Shukla, K., Inouye, M., and Montelione, G.T. 2003. Solution NMR structure of ribosome binding factor a (RbfA), a coldshock adaptation protein from *Escherichia coli*. *J. Mol. Biol.* **327:** 521–536.

Juszewski, K., Schwieters, C.D.S., Garrett, D.S., Byrd, R.A., Tjandra, N., and Clore, G.M. 2004. Completely automated, highly error-tolerant macromolecular structure determination from multi-dimensional nuclear Overhauser enhancement spectra and chemical shift assignments. *J. Am. Chem. Soc.* **126:** 6258–6273.

Kay, L.E., Clore, G.M., Bax, A., and Gronenborn, A.M. 1990. Four-dimensional heteronuclear triple-resonance nmr spectroscopy of interleukin-1β in solution. *Science* **249:** 364–365.

King, G.F., Shih, Y.L., Maciejewski, M.W., Bains, N.P., Pan, B., Rowland, S.L., Mullen, G.P., and Rothfield, L.I. 2000. Structural basis for the topological specificity function of MinE. *Nat. Struct. Biol.* **7:** 1013–1017.

Nilges, M. 1993. A calculation strategy for the structure determination of symmetric dimers by ^{1}H NMR. *Proteins* **17:** 297–309.

Nilges, M., Macais, M., Odonoghue, S., and Oschkinat, H. 1997. Automated NOESY interpretation with ambiguous distance restraints: The refined NMR solution structure of the pleckstrin homology domain from β-spectrin. *J. Mol. Biol.* **269:** 408–422.

O'Donoghue, S.I., Junius, F.K., and King, G.F. 1993. Determination of the structure of symmetric coiled-coil proteins from NMR data: Application of the leucine zipper proteins Jun and GCN4. *Protein Eng.* **6:** 557–564.

O'Donoghue, S.I., Chang, X., Abseher, R., Nilges, M., and Led, J.J. 2000. Unraveling the symmetry ambiguity in a hexamer: Calculation of the R6 human insulin structure. *J. Biomol. NMR* **16:** 93–108.

Oxenoid, K. and Chou, J.J. 2005. The structure of phospholamban pentamer reveals a channel-like architecture in membranes. *Proc. Natl. Acad. Sci.* **102:** 10870–10875.

Potluri, S., Yan, A.K., Chou, J.J., Donald, B.R., and Bailey-Kellogg, C. 2006. Structure determination of symmetric protein complexes by a complete search of symmetry configuration space using NMR distance restraints and van der Waals packing. *Proteins* **65:** 203–219.

Revington, M., Semesi, A., Yee, A., and Shaw, G.S. 2005. Solution structure of the *Escherichia coli* protein ydhR: A putative mono-oxygenase. *Protein Sci.* **14:** 3115–3120.

Seavey, B.R., Farr, E.A., Westler, W.M., and Markley, J. 1991. A relational database for sequence-specific protein NMR data. *J. Biomol. NMR* **1:** 217 236.

Walters, K.J., Matsuo, H., and Wagner, G. 1997. A simple method to distinguish intermonomer nuclear Overhauser effects in homodimeric proteins with C_2 symmetry. *J. Am. Chem. Soc.* **119:** 5958–5959.

Wang, L. and Donald, B.R. 2005. An efficient and accurate algorithm for assigning nuclear Overhauser effect restraints using a rotamer library ensemble and residual dipolar couplings. In *Proc IEEE Comput. Syst Bioinform Conf. (CSB2005)* pp. 189–202.

Wang, C.E., Pérez, T.L., and Tidor, B. 1998. AMBIPACK: A systematic algorithm for packing of macromolecular structures with ambiguous distance constraints. *Proteins* **32:** 26–42.

Wiltscheck, R., Kammerer, R.A., Dames, S.A., Schulthess, T., Blommers, M.J., Engel, J., and Alexandrescu, A.T. 1997. Heteronuclear NMR assignments and secondary structure of the coiled coil trimerization domain from cartilage matrix protein in oxidized and reduced forms. *Protein Sci.* **6:** 1734 1745.

Xu, Y., Wu, J., Gorenstein, D., and Braun, W. 1999. Automated 3D assignment and structure calculation of crambin (S22/I25) with the self-correcting distance geometry based NOAH/DIAMOD programs. *J. Magn. Reson.* **136:** 76–85.

Zwahlen, C., Legault, P., Vincent, S.J.F., Greenblatt, J., Konrat, R., and Kay, L.E. 1997. Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage λ N-Peptide/*boxB* RNA complex. *J. Am. Chem. Soc.* **119:** 6711–6721.